RESEARCH ARTICLE

BioDiscovery

OPEN ACCESS

Blood-Based Gene Expression in children with Autism spectrum disorder

Hristo Y. Ivanov[1*], Vili K. Stoyanova[1,2], Nikolay T. Popov[3], M. Bosheva[1,2], Tihomir I. Vachev[1,4]

[1]Department of Pediatrics and Medical Genetics, Medical University, Plovdiv Vasil Aprilov 15-A Str., Plovdiv 4002, Bulgaria
[2]University Hospital "St. George" Plovdiv, 66 Peshtersko Shuse Str., Plovdiv 4000, Bulgaria
[3]Psychiatric ward for active treatment of men, State Phsychiatry Hospital Pazardzhik, 28 Bolnichna Str. 4400 Pazardzhik, Bulgaria
[4]Department of Plant Phisyology and Molecular Biology, University of Plovdiv "Paisii Hilendarski", 24 Tzar Assen Str., Plovdiv, Bulgaria

Abstract

Comparative gene expression profiling analysis is useful in discovering differentially expressed genes associated with various diseases, including mental disorders. Autism spectrum disorder (ASD) is a severe neuropsychiatric disorder which has complex pathobiology with profound influences of genetic factors in its development. Although numerous autism susceptible genes were identified, the etiology of autism is not fully explained. The study aimed to examine gene expression profiling in peripheral blood from 60 individuals divided into two groups: children with ASD and age- and gender-matched healthy subjects (ASD control). A genome-wide sequencing of copy DNA molecules was conducted to obtain information for quantitative expression of all genes subject to this study and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway-based analysis was used to further understand genes' biological functions. Based on the conducted expression analysis 23 differentially expressed genes and 21 KEGG signaling pathways with statistical significant change were identified. Blood-based comparative gene expression profiling analysis is useful in discovering genetic markers associated with ASD. Our data will provide a valuable resource for discovery purposes and for comparison to other gene expression-based, genome-wide studies and other functional data.

Citation: Ivanov HY, Stoyanova VK, Popov NT, Bosheva M, Vachev TI. Blood-based gene expression in children with Autism spectrum disorder. *Biodiscovery* 2015; **17**: 2; DOI: 10.7750/BioDiscovery.2015.17.2

Copyright: © 2015 Ivanov et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, provided the original authors and source are credited.

Received: 26 August 2015; **Accepted:** 28 September 2015; **Available online/Published:** 30 September 2015

Keywords: ASD, transcriptome sequencing, gene expression, gene pathway analysis

*Corresponding Author: Hristo Yordanov Ivanov, email: doctorhristoivanov@yahoo.com

Conflict of Interests: No potential conflict of interest was disclosed by any of the authors.

INTRODUCTION

Autism spectrum disorder (ASD) is an entity that reflects a scientific consensus concerning several previously separated disorders, actually based on current definition a single spectrum disorder with different levels of symptom severity in two core domains - deficits in social communication and interaction, and restricted repetitive behaviors [1]. According to the Center for Disease Control (CDC) and Autism and Developmental Disabilities Monitoring (ADDM) Network in the U.S. 1 in 88 children are identified with the ASD [2] although this increase might merely reflect improved diagnostic means. The etiology of ASD is mainly attributable to genetic factors. The estimated heritability of ASD is more than 90%, and the genetic basis of ASD are heterogeneous and complex, involving multiple genes, gene-gene interactions, and gene-environmental interactions [3]. Identification of genetic basis can shed a light on etiology and pathogenesis of this disorder which still remain elusive. The genetic risk factors for ASD identified so far range from common variants conferring a small clinical effect to rare mutations that confer a high clinical effect [4]. Several genes have been found to be associated with ASD, but more genes

remain to be discovered.

Microarray-based gene expression profiling is a useful technology allowing simultaneous measurement of hundreds to thousands of gene transcripts, which is useful for large-scale gene discovery studies [5]. This technology has been used in several post-mortem brain studies of psychiatric disorders, including schizophrenia, bipolar disorder, and autism [6]. Since there are several limitations to the use of the post-mortem brain tissue in gene expression studies, and the study of fresh brain tissue from living psychiatric patients is impractical at the present time, several studies have reported the use of peripheral blood cells and lymphoblastoid cell lines (LCL) as surrogates for brain tissue. [7]. In addition, a moderate correlation of gene expression between peripheral blood cells and brain tissue in humans has been reported, supporting the usefulness of peripheral blood cells in the gene expression studies for psychiatric research [8].

The above-mentioned findings attracted our interest in identifying novel differentially expressed genes associated with ASD in our population using comparative gene expression profiling based on the next generation sequencing analysis of peripheral blood samples.

RESULTS

Based on the conducted expression analysis we identified 23 differentially expressed genes with statistical significant change ($p < 0.01$, FDR < 0.001 and fold change > 1), (Table 1).

Table 1. Differentially expressed genes with statistically significant changes.

GeneID	Symbol	Description	P-value	FDR	Fold Change
6543	SLC8A2	solute carrier family 8 (sodium/calcium exchanger), member 2	2.7869e-10	2.5e-08	5.18
118932	ANKRD22	ankyrin repeat domain 22	2.7132e-11	2.9e-09	1.78
51282	SCAND1	SCAN domain containing 1	9.2933e-06	0.00036382	1.66
5909	RAP1GAP	RAP1 GTPase activating protein	6.07386e-10	5.2e-08	1.43
54344	DPM3	dolichyl-phosphate mannosyltransferase polypeptide 3	1.31106e-05	0.0004975	1.23
286204	CRB2	crumbs homolog 2 (Drosophila)	3.15268e-06	0.0001	1.11
22974	TPX2	TPX2, microtubule-associated, homolog (Xenopus laevis)	2.43966e-07	1.3e-05	1.10
3118	HLA-DQA2	major histocompatibility complex, class II, DQ alpha 2	3.29222e-11	3.5e-09	1.06
57126	CD177	CD177 molecule	2.71846e-07	1.4e-05	1.05
4057	LTF	lactotransferrin	1.53388e-16	2.4e-14	1.00
4071	TM4SF1	transmembrane 4 L six family member 1	5.80188e-11	5.7e-09	-7.31
4103	MAGEA4	melanoma antigen family A, 4	1.07706e-05	0.0004	-6.26
2318	FLNC	filamin C, gamma	1.34649e-13	1.7e-11	-5.22
10642	IGF2BP1	insulin-like growth factor 2 mRNA binding protein 1	9.62384e-08	5.7e-06	-4.41
8061	FOSL1	FOS-like antigen 1	1.38726e-08	9.9e-07	-2.13
9381	OTOF	otoferlin	2.35062e-05	0.0008	-1.76
7503	XIST	X (inactive)-specific transcript (non-protein coding)	4.07256e-16	6.24e-14	-1.38
51127	TRIM17	tripartite motif containing 17	1.14521e-06	5.3e-05	-1.31
55287	TMEM40	transmembrane protein 40	2.79288e-08	1.8e-06	-1.22
5054	SERPINE1	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1	4.7984e-07	2.4e-05	-1.20
54841	BIVM	basic, immunoglobulin-like variable motif containing	1.9136e-05	0.000687	-1.13
618	BCYRN1	brain cytoplasmic RNA 1 (non-protein coding)	4.79516e-06	0.0001	-1.07
653160	LOC653160	hypothetical LOC653160	1.38538e-05	0.0005	-1.02

KEGG pathway enrichment analisys of DEGs

Genes usually interact with each other to play roles in certain biological functions. Pathway-based analysis helps to further understand genes biological functions. We use the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database, which is the major public pathway-related database, [9] to extract the relations between ASD and the affected biological pathways. Pathway enrichment analysis identifies significantly enriched metabolic pathways and signal transduction pathways in DEGs comparing with the whole genome background. We carried out a pathway-based analysis to identify the risk pathways of ASD. The most significant 21 KEGG signaling pathways for both up- and down regulated gene sets in the significance order (size of P values) are listed in Table 2.

DISSCUSSION

The data show differences in gene expression in whole blood of children with autism compared to typically developing children matched for age and gender to the general population. Whole-transcriptome sequencing of pools of all children with autism compared to controls gives comprehensive information on a large number of significantly differentially expressed genes, suggesting that some aspect of "autism" is shared by most of the children sampled. Pathway analyses based on nominally significant genes revealed 21 KEGG pathways, significantly changed in ASD children compared to controls (nominal P-value < 0.05), involved in signaling, immune system and metabolic processes. Our findings suggest that synaptic plasticity, disturbances in axon guidance, neuronal survival, differentiation and inflammation have great role in the pathophysiology of ASD. We briefly discuss some of the most significant pathways below.

Calcium signaling pathway

Dysregulation of calcium signaling pathway is interesting since recent collaborative effort of meta-analysis of five major neuropsychiatric disorders including ASD suggests that two calcium channel coding genes – CACNA1C and CACNB2 – are significantly associated with all five

Table 2. Pathway enrichment analysis of DEGs.

#	Pathway	DEGs with pathway annotation (17)	All genes with pathway annotation (16502)	P value	Pathway ID
1	Calcium signaling pathway	3 (17.65%)	272 (1.65%)	0.002537809	ko04020
2	Amphetamine addiction	2 (11.76%)	130 (0.79%)	0.007751596	ko05031
3	Leishmaniasis	2 (11.76%)	132 (0.8%)	0.007983234	ko05140
4	GABAergic synapse	2 (11.76%)	133 (0.81%)	0.008100223	ko04727
5	Retrograde endocannabinoid signaling	2 (11.76%)	137 (0.83%)	0.008575949	ko04723
6	MAPK signaling pathway	3 (17.65%)	430 (2.61%)	0.009102987	ko04010
7	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	2 (11.76%)	163 (0.99%)	0.01196498	ko05412
8	Cholinergic synapse	2 (11.76%)	167 (1.01%)	0.01253102	ko04725
9	GnRH signaling pathway	2 (11.76%)	184 (1.12%)	0.01506537	ko04912
10	Glutamatergic synapse	2 (11.76%)	187 (1.13%)	0.01553389	ko04724
11	Serotonergic synapse	2 (11.76%)	203 (1.23%)	0.01813789	ko04726
12	Dopaminergic synapse	2 (11.76%)	206 (1.25%)	0.01864557	ko04728
13	Steroid biosynthesis	1 (5.88%)	29 (0.18%)	0.02947291	ko00100
14	Wnt signaling pathway	2 (11.76%)	270 (1.64%)	0.0308568	ko04310
15	Leukocyte transendothelial migration	2 (11.76%)	296 (1.79%)	0.03652306	ko04670
16	Hypertrophic cardiomyopathy (HCM)	2 (11.76%)	321 (1.95%)	0.04232472	ko05410
17	Dilated cardiomyopathy	2 (11.76%)	330 (2%)	0.04449439	ko05414
18	Asthma	1 (5.88%)	45 (0.27%)	0.04538188	ko05310
19	Influenza A	2 (11.76%)	334 (2.02%)	0.04547207	ko05164
20	Carbohydrate digestion and absorption	2 (11.76%)	348 (2.11%)	0.04895749	ko04973
21	Endocrine and other factor-regulated calcium reabsorption	2 (11.76%)	349 (2.11%)	0.04921018	ko04961

diseases [10].

Calcium signals culminate by initiating the fusion of synaptic vesicles into the neuronal presynaptic membrane, and the sum of the evidence is that pre- and postsynaptic membrane changes recognized in several fundamental processes of synaptic plasticity and learning are calcium sensitive and are altered in models of ASD. Recent findings in the etiopatogenesis of the ASD-causing Angelman, Prader-Willi, Rett, **tuberous sclerosis** suggest an inability of neurons to generate adaptive responses via calcium-regulated gene expression [11].

MAPK signaling pathway

Dysregulation in Ras/mitogen-activated protein kinase (Ras/MAPK) pathway genes lead to a class of disorders known as RASopathies which includes neurofibromatosis type 1 (NF1), Costello syndrome (CS), Noonan syndrome (NS), and cardio-facio-cutaneous syndrome (CFC). Potential genetic and phenotypic overlap between dysregulation of Ras/MAPK signalling and ASD were suggested in previous studies. Higher prevalence and severity of ASD traits in RASopathies compared to unaffected siblings propose that dysregulation of Ras/MAPK signalling during development may be implicated in ASD risk [12].

Wnt signaling pathway

The Wnt pathway is involved in various well-known cellular processes including differentiation and migration, especially during nervous system development and cell proliferation. Given its various functions, dysfunction of the canonical Wnt pathway probably exert adverse effects on neurodevelopment and therefore leads to the pathogenesis of autism.

In the last years an increased amount of evidence has shown that components of Wnt pathways are involved in major psychiatric disorders. A literature review supports the contention that modification of genes affecting the activity of the Wnt pathway could contribute to individual forms of ASD [13].

Glutamatergic and GABAergic synapse

Critical balance between excitatory glutamate and inhibitory GABA neurotransmitter is essential and crucial for proper development and functioning of brain. GABAergic (gamma aminobutyric acid) and glutamatergic interneurons maintain excitability, integrity and synaptic plasticity. Several recent evidences implicated relative loss of inhibitory GABA with corresponding glutamate mediated hyper-excitation in development of ASD. Moreover, several studies have demonstrated the imbalance of excitatory/inhibitory neurotransmitters resulting from neurodevelopmental impairments in glutamatergic and GABAergic system, which might resemble a common pathological mechanism underlying developmental disorders [14].

In summary, our list of differentially regulated genes is enriched with pathways associated with nervous system development and function, and immune system and most of them seem to be around core networks such as those involved in kinase and/or signaling networks. Therefore, our results support the involvement of various genetic factors (heterogeneity) in the development of ASD, while suggesting these different factors can be converging at, or diverging from central networks such as signaling networks.

Findings from our current study demonstrate that there are clear and significant abnormalities in the gene expression of peripheral blood samples obtained from children with ASD compared to healthy controls. This promising work, while far from being definitive, gives further proof to the recently emerging principle that peripheral blood is a potentially useful source of diagnostic biomarkers for disorders of the brain and other inaccessible tissues [15]. If the results of this work are confirmed in future studies and the identified changes in the study group are individually validated by us or by others in other independent cohorts, we can assume that the differentially expressed genes may help clarify the etiology and pathogenesis of this disorder and the dysregulated pathways may also provide targets for the experimental treatments in ASD.

MATERIALS AND METHODS

Subjects

A total of 30 subjects (24 - male; 6 - female) with idiopathic ASD aged 3 to 11 years (mean age of the sample - 6.86 years), and 30 healthy children age and sex matched to the patient's group were recruited for gene expression analysis Participants were interviewed by experienced child psychiatrists. The diagnose ASD was established by the use of ADI-R [16], CARS or GARS [17]. All of the participants were Bulgarians. None of the participants had received any medications before blood sampling.

Written informed consent and child assent were obtained from the parents and the probands, respectively, after the purposes and procedures of the study were fully explained and confidentiality was ensured.

Laboratory methods

Blood collection and RNA isolation

An aliquot of peripheral whole blood samples (2.5 ml) for each subject (ASD diagnosed and healthy controls) was collected into PAXgene blood RNA tubes (PreAnalytiX) and stored for minimum of 4 hours and then freezed at −80°C, as this method shows the biggest yield of RNA [18]. After collection of all samples, total RNA was isolated using the PAXgene blood RNA kit (PreAnalytiX), according to the manufacturer's protocol. The resulting total RNA samples was treated with RNase-free DNase

I (Promega) according to the manufacturer's protocol followed by PCR checking for DNA contaminations.

Quantitative analysis of the isolated total RNA samples

The measured absorbance ratio of 260 nm and 280 nm (A260/A280) was between 1.93 and 2.1 for all samples included for further analysis, according to the service requirements. Pooled samples were created by adding an equivalent amount of total RNA from each individual sample to final concentration of 5 µg RNA samples. Pooled RNA samples were precipitated according to the service requirements, each pooled RNA sample was mixed with 1/10th volume of 3M NaOAc, pH 5.2 and 3 volume 100% ethanol, to the final volume of 400 µl. Aliquots of pooled RNAs were frozen at -80°C and shiped on dry ice. RNA integrity of pooled samples (ASD and control group) was assessed by agarose gel electrophoresis (Fig. 1) and checked by Agilent 2100 Bioanalyzer (Fig. 2 and 3).

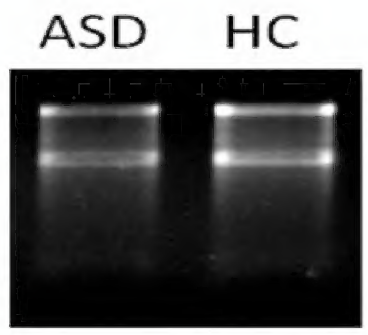

Figure 1. Agarose gel analysis of the pooled RNA samples (Pooled samples from ASD group (ASD) and healthy controls (HC), respectively.

RNA-Seq (quantification) analysis

In this study we used Beijing Genomics Institute (BGI) as a Certified Service Provider for sequencing service. The total RNA samples were first treated with DNase I to degrade any possible DNA contamination. Then the mRNA was enriched by using the oligo(dT) magnetic beads. Mixed with the fragmentation buffer, the mRNA was fragmented into short fragments (about 200 bp). Then the first strand of cDNA was synthesized by using random hexamer-primer. Buffer, dNTPs, RNase H and DNA polymerase I were added to synthesize the second strand. The double strand cDNA was purified with magnetic beads. End reparation and 3'-end single nucleotide A (adenine) addition was then performed. Finally, sequencing adaptors were ligated to the fragments. The fragments were enriched by PCR amplification. During the QC step, Agilent 2100 Bioanaylzer (Fig.1 and 2). and ABI StepOnePlus Real-Time PCR System were used to qualify and quantify of the sample library. The library products were sequencing via Illumina HiSeqTM 2000.

Quantification of gene expression

The expression level for each gene was determined by the numbers of reads uniquely mapped to the specific gene and the total number of uniquely mapped reads in the sample. The gene expression level was calculated by using RPKM. [19] (Reads Per Kilobase per Million

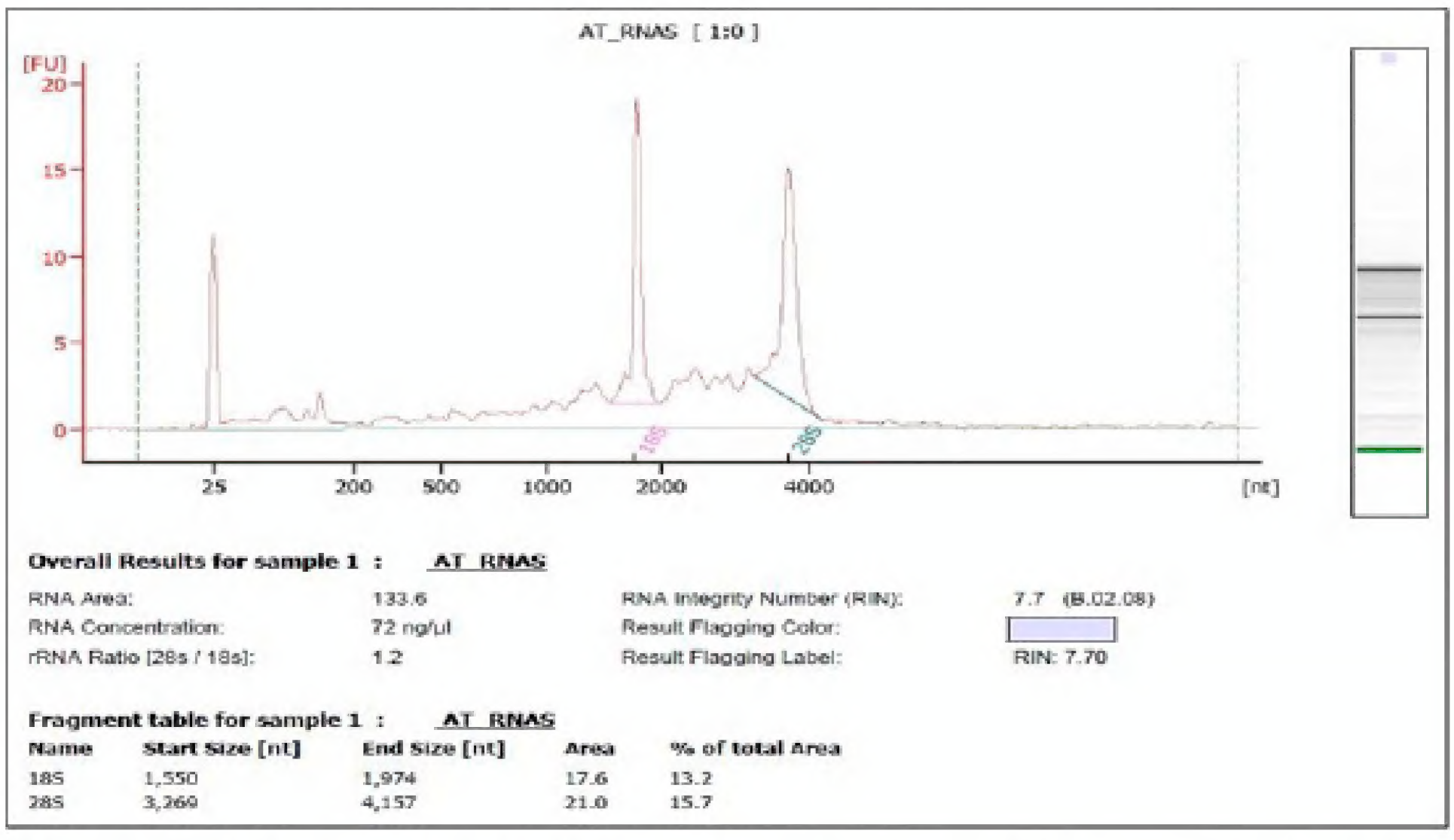

Figure 2. Agilent 2100 Bioanalyzer results of the pooled RNA sample from ASD group (ASD).The two tall peaks at 2000 and 4000 are the ribosomal RNA. RNA integrity number (RIN) of the pooled total RNA were calculated from the bioanalyzer traces. Pooled total RNA has RNA integrity number (RIN) 7.7.

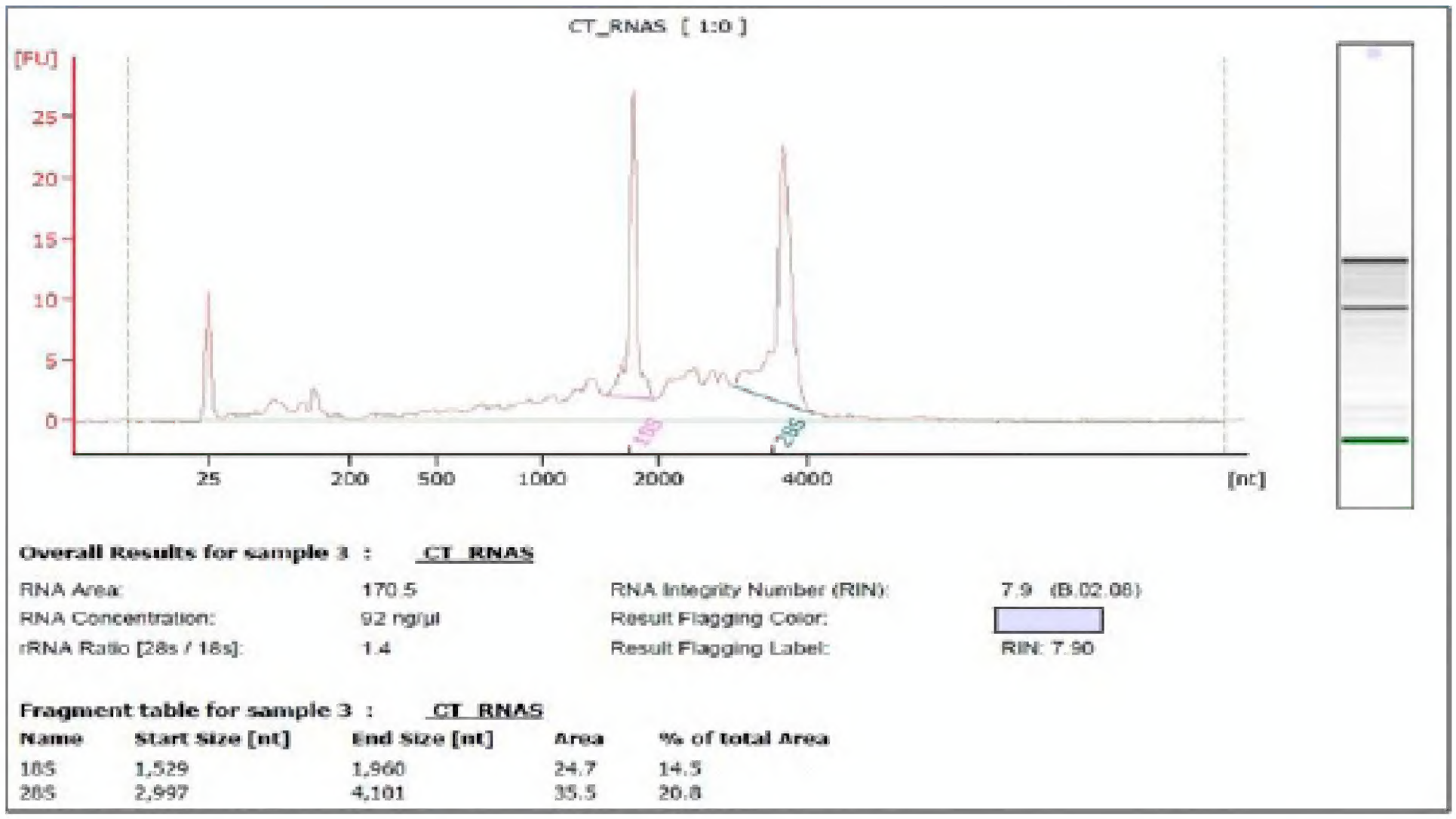

Figure 3. Agilent 2100 Bioanalyzer results of the pooled RNA sample from the healthy control group (HT). The two tall peaks at 2000 and 4000 are the ribosomal RNA, RNA integrity number (RIN) of the pooled total RNA were calculated from the bioanalyzer traces. Pooled total RNA has RNA integrity number (RIN) 7.9.

mapped reads) method, and the formula is shown as follows:

$$RPKM = \frac{10^6 C}{NL/10^3}$$

RPKM is a method of quantifying gene expression from RNA sequencing data by normalizing for total read length and the number of sequencing reads. The RPKM method is able to eliminate the influence of different gene length and sequencing discrepancy on the calculation of gene expression level. Therefore, the RPKM values can be directly used for comparing the difference of gene expression among samples. If there is more than one transcript for a gene, the longest one is used to calculate its expression level and coverage.

Screening of differentially expressed genes (DEGs)

This analysis includes the screening of genes that are differentially expressed among samples and KEGG pathway enrichment analysis for these DEGs. Referring to "The significance of digital gene expression profiles" [20], a strict algorithm to identify differentially expressed genes between two samples have been developed.

Denote the number of unambiguous clean tags from gene A as x, given every gene's expression occupies only a small part of the library, p(x) will closely follow the Poisson distribution.

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$ (λ is the real transcripts of the gene)

The total clean tag number of the sample 1 is N1, and total clean tag number of sample 2 is N2; gene A holds x tags in sample 1 and y tags in sample 2. The probability of gene A expressed equally between two samples can be calculated with:

$$2\sum_{i=0}^{i=y} p(i \mid x)$$

$$\text{or} \quad 2\times(1-\sum_{i=0}^{i=y} p(i \mid x)) \qquad (\text{if} \sum_{i=0}^{i=y} p(i \mid x) > 0.5)$$

$$p(y \mid x) = (\frac{N_2}{N_1})^y \frac{(x+y)!}{x!\,y!(1+\frac{N_2}{N_1})^{(x+y+1)}}$$

P-value corresponds to differential gene expression test. FDR (False Discovery Rate) is a method to determine the threshold of P-value in multiple tests and assume that we have picked out R differentially expressed genes in which S genes really show differential expression and the other V genes are false positive. If we decide that the error ratio „Q = V / R" must stay below a cutoff (e.g. 1%), we should preset the FDR to a number no larger than 0.01 [21]. „FDR ≤ 0.001 and the absolute value of log2Ratio ≥ 1" as the threshold was used to judge the significance of gene expression difference.

References

1. A.P. Association, The Diagnostic and Statistical Manual of Mental Disorders: DSM 5, 2013.
2. Baio J. Prevalence of Autism Spectrum Disorders - Autism and Developmental Disabilities Monitoring Network, 14 Sites, United States, 2008 Surveillance Summaries March 30, 2012 / 61(SS03);1-19.
3. Eapen V: Genetic basis of autism: is there a way forward? *Curr Opin Psychiatry* 2011; **24(3):** 226-236.
4. State MW, Levitt P: The conundrums of understanding genetic risks for autism spectrum disorders. *Nat Neurosci* 2011; **14(12):** 1499-1506.
5. Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci USA* 1996, **93(20):** 10614-10619.
6. Konradi C: Gene expression microarray studies in polygenic psychiatric disorders: applications and data analysis. *Brain Res Brain Res Rev* 2005; **50(1):** 142–155.
7. Woelk CH, Singhania A, Pérez-Santiago J, Glatt SJ, Tsuang MT. The utility of gene expression in blood cells for diagnosing neuropsychiatric disorders. *Int Rev Neurobiol* 2011; **101:** 41–63.
8. Sullivan PF, Fan C, Perou CM. Evaluating the comparability of gene expression in blood and brain. *Am J Med Genet B Neuropsychiatr Genet* 2006; **141B(3):** 261–268.
9. Kanehisa, M., Araki M, Goto S, Hattori M, Hirakawa M, Itoh M et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res* 2008; **36** (Database issue): D480-4.
10. Cross-Disorder Group of the Psychiatric Genomics, Consorcium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 2013; **381(9875):** 1371-1379.
11. J. Jay Gargus, Galina Schmunk Dysregulation of Neurogenic Calcium Signaling and Autism in Comprehensive Guide to Autism, *Springer New York* 2014, pp 1285-1312.
12. Adviento B, Corbin IL, Widjaja F, Desachy G, Enrique N, Rosser T, et al. Autism traits in the RASopathies. *J Med Genet.* 2014; **51(1):** 10-20.
13. Zhang Y, Yuan X, Wang Z, Li R: The Canonical Wnt Signaling Pathway in Autism. *CNS Neurol Disord Drug Targets* 2014; **13(5):** 765-770.
14. Purkayastha P, Malapati A, Yogeeswari P, Sriram DA. Review on GABA/Glutamate Pathway for Therapeutic Intervention of ASD and ADHD. *Curr Med Chem* 2015 Feb 9[Epub ahead of print]
15. Kohane IS, Valtchinov VI. Quantifying the white blood cell transcriptome as an accessible window to the multiorga n transcriptome. *Bioinformatics*. 2012; **28(4):** 538-545.
16. Lord C, Rutter M, Couteur AL. Autism Diagnostic Interview–Revised: A Revised Version of the Diagnostic Interview for Caregivers of Individuals with Possible Pervasive Developmental Disorders. *J Autism Dev Disord* 1994; **24(5):** 659-685.
17. Gilliam, J. Gilliam Autism Rating Scale Examiner‘s Manual. Austin, 1995; TX: Pro-Ed.
18. T. Vachev, I Minkov, V Stoyanova, N Popov. Down regulation of miRNA let-7b-3p and let-7d-3p in the peripheral blood of children with autism spectrum disorder. *Int.J.Curr.Microbiol. App.Sci* 2013; **2(12):** 384-388.
19. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* 2008; **5(7):** 621-628.
20. Audic S, Claverie JM. The significance of digital gene expression profiles. *Genome Res* 1997; **7(10):** 986-995.
21. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Stat.* 2001; **29(4):** 1165-1188.